Neural Network-based Blood Glucose Control System for Type I Diabetes Patients

Serge Chernetsov[1], Anatoly Karpenko[2], Alexander Trofimov[3]

[1,2]Computer-Aided Design Department, Moscow State Technical University n.a. Bauman
5, 2nd Baumanskaya St., Moscow, Russia

[3]Cybernetics Departament, National Nuclear Research University (MEPhI)
31, Kashirskoye h/w, Moscow, Russia

[1]serge.a.ch@gmail.com; [2]apkarpenko@mail.ru; [3]atrofimov@list.ru

***Abstract*- The paper considers the problem of constructing neural network-based blood glucose control system, consisting of connected neural networks of different architectures, NARX neural networks and TDNN. The principle underlying the model training and the results of experimental studies on real patient data are described.**

Keywords- Neural Network; Blood Glucose Control System; Type I Diabetes; IDDM; Control System

I. INTRODUCTION

At the core of pathogenetic mechanism, type I diabetes lays cells lack of insulin synthesis and secretion of the pancreas endocrine caused by their destruction as a result of the impact of various factors. Consequently, impaired carbohydrate metabolism of body cells, in the absence of adequate treatment can lead to severe complications. The main treatment method of diabetes mellitus type Ⅰ, which is also insulin-dependent diabetes mellitus (IDDM), is insulin injections to normalize body metabolism. The purpose of injecting insulin is to control blood glucose and keep it in the physiological range.

The optimal insulin dose depends on many factors and is specific to each patient. Finding such doses is a challenge, not all patients can cope with. To help solute this problem, continuous glucose monitoring systems (CGMS) and continuous subcutaneous insulin infusion systems (insulin pumps) were created. Using these aids, automatic blood glucose control systems are being extensively developed.

These control systems are designed as a device, predicting the required dose of insulin at every given moment. The main problem in constructing such devices is to develop adequate algorithms, which would take into account the individual characteristics of glucose and insulin metabolism of patient's cells as well as many other factors that have a significant impact on the concentration of these substances in the blood.

A necessary step in building control systems is to build the object model, which is the mechanism of glucose processing by the cells in the body under the action of insulin. Despite the extraordinary complexity of the mechanism to date, there are a number of simplified analytical models describing the insulin and glucose dynamics in the blood (Bergman model, Staris, Nikita, Engelborga, etc.). Typically, these models contain a number of empirically estimated parameters for each individual patient. A detailed review of existing glucose metabolism models is given in [1].

The high complexity of the mechanism of insulin and glucose metabolism, the individuality of its characteristics, as well as the lack of accurate mathematical models and rules for calculating the required dose of insulin to ensure the maintenance of glucose in the physiological range, are the factors which make use of the methodology of soft computing, artificial neural networks in particular, very efficient in constructing the blood glucose control system [1].

It is known that neural networks with high computational capabilities can be trained on experimental data and provide substantial benefit performance-wise, which allows building robust models capable of taking into account peculiarities of each individual patient [3]. Furthermore, dynamic neural networks show good results in dynamic object modelling and in control systems [3]-[7].

In this paper, neural network-based models of the glucose and insulin dynamics, based on the use of neural networks with time delays – nonlinear auto regressive exogenous (NARX), neural networks, and time-delay neural networks (TDNN) – is considered [2],[3]. The model class choice is made in favour of NARX networks due to their high generalization abilities [3],[12],[13]. Using simpler TDNN as a control system simplifies the training procedure [14].

However, other model classes can be used as an object model, e.g. the models considered in [15] and [16] providing precision compatible to that presented in this research. But after comparison of that with more traditional forecasting methods [17], the neural networks are in favour.

II. PROBLEM STATEMENT

The initial data for constructing a model of the insulin and glucose dynamics are the blood glucose readings, carbohydrate intake and the insulin dose up to the moment. Note that this data set is probably not sufficient for an accurate model of the glucose dynamics – blood glucose levels are also affected by many other external and internal factors. However, the developed model can be easily modified taking into account a number of additional external factors.

Suppose that all observations are grouped in the P series, each of which represents a set of discrete observations for a certain period of time, conducts on a regular basis at regular intervals. Denote, $\sigma^{(p)}(t)$, $\chi^{(p)}(t)$ and $x^{(p)}(t)$ are the observed blood glucose, insulin dose and the values of external and internal factors affecting the metabolism of glucose and insulin, respectively. In discrete time t in the pth series of observations, $t = \overline{1, T_p}$ where Tp – is the number of observations in the p series $p = \overline{1, P}$. Note that the value of $x^{(p)}(t)$, which is responsible for internal and external factors, in general, can be a vector. In this paper, this value represents the amount of

carbohydrates consumed. Furthermore, for simplicity, the superscript (p) will be omitted wherever the number of series of observations is not important.

In this paper we consider two schemes of control: open-loop (Fig. 1) and closed loop (Fig. 2). The object is the mechanism of glucose metabolism in the cells of the body, the input to which at time t is the insulin dose χ(t), defined by a medical expert or by the patient (reference dose); the output is the blood glucose reading σ(t). Denote as y(t) glucose levels observed at the output of the object model at time t, and the simulated dose of insulin u(t).

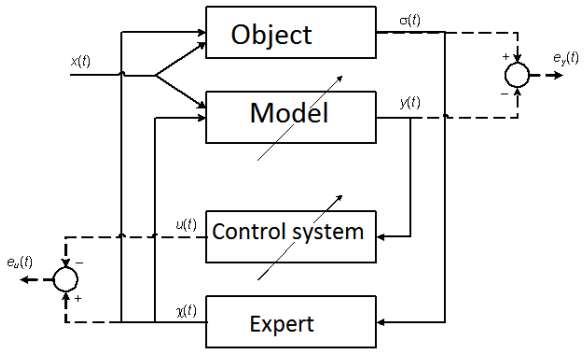

Fig. 1 Open-loop control system

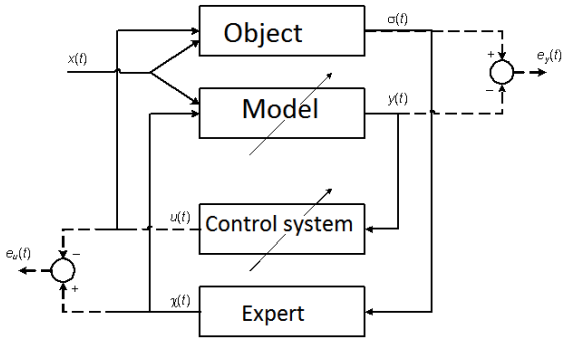

Fig. 2 Closed-loop control system

The available source data sets $\left\{\chi^{(p)}(t), t=\overline{1,T_p}\right\}$ and $\left\{\sigma^{(p)}(t), t=\overline{1,T_p}\right\}$ from pth series, can be regarded as the result of the functioning of the outer loop on the control schemes presented. In an open-loop circuit at the input of the control object model at a moment of time, a reference dose of insulin χ(p)(t) is fed, while in the closed circuit the calculated insulin dose u(p)(t) is applied to the input of the controlled object model. Mismatch $e_u^{(p)}(t)=\chi^{(p)}(t)-u^{(p)}(t)$ and $e_y^{(p)}(t)=\sigma^{(p)}(t)-y^{(p)}(t)$, between the reference and modelled values of glucose and insulin are used to fine-tune the object model and control system.

The closed-loop circuit contains two independent control loops: outer defining input data for modelling, and interior formed by adaptive models. In this scheme, the output of the outer loop is not directly involved in the models of the inner loop.

In both diagrams of the process of training, the model learns to simulate the behaviour of the object, and control device, which is actions of the expert. Outside the experiment, the correction by expert will be absent and the closed-loop system will operate only inside the control loop, while in the open-loop mode of operation of the control system must be closed. Due to the fact that the adaptation of this system was carried out in the open mode, and is supposed to use in a closed mode, accumulation of modeling errors is possible in such a system. This problem does not exist in closed-loop control system.

Taking into account the high generalizing ability of NARX neural network models in the simulation of dynamic systems [2], it is this class of neural network models who has been chosen as an object model. Neural network-based NARX model consists of a multilayer perceptron (MLP), which is fed with the input of an external control signals through the tapped delay lines (TDL) and its own output through the feedback and the TDL. The results of NARX neural networks to predict the level of glucose in the blood of patients with diabetes mellitus type 1 are given in [10], [11].

The task of the control system is to calculate the required insulin dose $u(t)$ at the moment t. We assume that the required insulin dose $u(t)$ is a function of current and previous blood glucose values:

$$u(t)=F\left(y(t),t(t-1),...,y(t-m)\right) \qquad (1)$$

This assumption is used in constructing the models, described in [11]. In more complex cases, it is assumed that the required insulin dose $u(t)$ explicitly depends on the insulin dose in previous points in time.

Given (1) we will use TDNN as a control system. Let us pose the problem of training the object model and the control system. Training will be conducted in such a way that minimize the mismatch between model outputs and the reference values both of the object model and of the control system on existing data samples. We will use the following training Criterion (2):

$$I=\frac{1}{2}\sum_{p=1}^{P}\sum_{t=t_0}^{T_p}\left(\alpha^{(p)}(t)\left(y^{(p)}(t)-\sigma^{(p)}(t)\right)^2+\beta^{(p)}(t)\left(u^{(p)}(t)-\chi^{(p)}(t)\right)^2\right) \qquad (2)$$

where $\alpha^{(p)}(t)$ and $\beta^{(p)}(t)$ are weights of p^{th} series at the moment t. The task of training neural network control system is to minimize this criterion by adjusting the model parameters: $I\to\min_{w,v}$ where w and v are adjustable parameters vectors of object model and control system respectively.

Initial insulin dose and glucose levels in the current series are chosen as the initial model states (output delay lines at the moment t_0).

III. CONTROL SYSTEM TRAINING PROCEDURE

Training neural network control system is minimizing the Criterion (2) by adjusting the parameters of multilayer perceptron, which are members of the NARX neural network and TDNN. To tune fine the elements of the vectors w and v, we will apply gradient methods using information about the partial derivatives of these parameters. To derive the partial derivatives equations, we use the principle underlying the error back propagation [2].

Operation of the considered closed-loop control systems in time on the pth series of observations can be represented as operation of "multilayer" network that has 2*(Tp-t0+1) "layers", where each odd layer which is a NARX neural network, and each even layer, TDNN, i.e. implement its unfolding in time [2]. A portion of the unfolded closed-loop

control system is shown in Fig. 3. Double arrow in the figure denotes the vector signal.

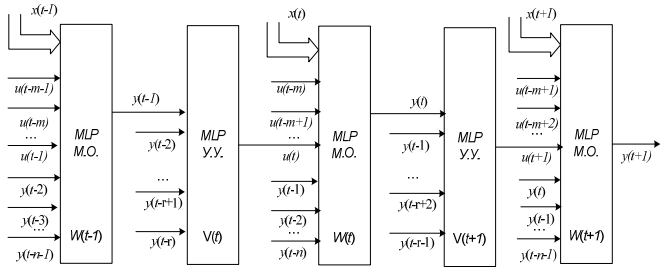

Fig. 3 Fragment of closed-loop control system unfolded in time

Denote $I^{(p)} = \frac{1}{2}\sum_{t=t_0}^{T_p}\alpha^{(p)}(t)e_y^{(p)}(t)^2 + \frac{1}{2}\sum_{t=t_0}^{T_p}\beta^{(p)}(t)e_u^{(p)}(t)^2$, in this case, training criterion is $I = \sum_{i=1}^{P} I^{(p)}$, and partial derivatives are $\frac{\partial I}{\partial \xi} = \sum_{p=1}^{P}\frac{\partial I^{(p)}}{\partial \xi}$, where ξ refers any element from vectors w and v.

In order to calculate $\frac{\partial I^{(p)}}{\partial \xi}$, consider the "remaining loss" function: $I^{(p)}(t) = \sum_{\tau=t}^{T_p} I_\tau^{(p)}$,

where $I_\tau^{(p)} = \frac{1}{2}\alpha^{(p)}(\tau)e_u^{(p)}(\tau)^2 + \frac{1}{2}\beta^{(p)}(\tau)e_y^{(p)}(\tau)^2$ refers current losses in the moment τ of network operation. In this case, training criterion equals $I = \sum_{p=1}^{P} I^{(p)}(t_0)$.

Denote synaptic weights in t^{th} instance of unfolded NARX and TDNN as w(t) and v(t) respectively.

Taking into account the fact that the criterion value I_τ depends neither on w(t) nor on v(t)

$\frac{\partial I^{(p)}}{\partial \xi} = \sum_{t=t_0}^{T_p}\frac{\partial I^{(p)}}{\partial \xi(t)} = \sum_{t=t_0}^{T_p}\frac{\partial I^{(p)}(t)}{\partial \xi(t)}$, where $\xi(t)$ refers any of the vectors w(t) and v(t).

To compute the equation for $\frac{\partial I^{(p)}(t)}{\partial \xi(t)}$, use the ratio, which is true for remaining loss function $I^{(p)}(t) = I_t^{(p)} + I^{(p)}(t+1)$. All mentioned above yields to the following ratio for NN outputs in t^{th} instance:

$$\frac{\partial I^{(p)}(t)}{\partial y(t)} = \frac{\partial I_t^{(p)}}{\partial y(t)} + \frac{\partial I^{(p)}(t+1)}{\partial y(t)}.$$

Calculations of $\frac{\partial I^{(p)}(t)}{\partial w(t)}$ and $\frac{\partial I^{(p)}(t)}{\partial v(t)}$ are not presented in this paper because of the equation size.

In order to train open-loop system, the following scheme is applied. Since at any given moment t, the reference value of insulin dose $\chi(t)$ is fed to the input of NARX neural network, which models the object. Input vector $\tilde{u}(t) = (x(t), \chi(t))$ may be considered as external, i.e. independent of control system operation at any given moment of time. So the following two-stage training procedure is proposed. In the first stage, NARX object model disconnected from control system is trained in accordance with the criterion

$$I_y = \frac{1}{2}\sum_{p=1}^{P}\sum_{t=t_0}^{T_p}\alpha^{(p)}(t)\left(y^{(p)}(t) - \sigma^{(p)}(t)\right)^2 \quad (3)$$

In the second stage, control system is trained according to the criterion

$$I_u = \frac{1}{2}\sum_{p=1}^{P}\sum_{t=t_0}^{T_p}\beta^{(p)}(t)\left(u^{(p)}(t) - \chi^{(p)}(t)\right)^2 \quad (4)$$

Training is conducted in accordance with the gradient method. Formulas for the calculation of partial derivatives of the Criterion (3) for NARX neural networks are known and presented, for instance in [2]. Training TDNN in accordance with the Criterion (4) is also carried out on well-known formulas given in [2].

IV. EXPERIMENT RESULTS

Input data for simulation were obtained using Medtronic CGMS and insulin pump which are the blood glucose readings. Insulin dose and the amount of absorbed carbohydrates took every ten minutes in a patient with diabetes mellitus type Ⅰ.

The number of observations series P = 1, the number of time samples T = 1040. Before the construction of neural network, models available data were divided into a training (first 80% of the data, 831 samples) and test (20% of the data, 209 samples) samples. Graphs of the original data are shown in Fig. 4.

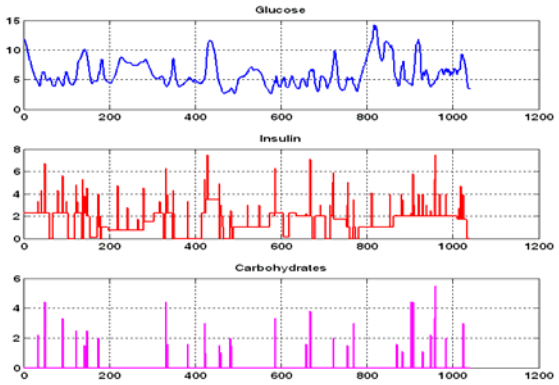

Fig. 4 Input data above: the blood glucose; in the center: insulin; below: the amount of carbohydrates consumed. The abscissa is the number of time samples

A. Simulation of Open-loop Control System

In the first stage of building, an open-loop control system NARX neural network is trained model in accordance with the Criterion (3). In the second phase, TDNN modeling the control device is trained in accordance with the Criterion (4). As a result, experimental studies found that better performance accuracy of the model on a test set, as defined in accordance with the Criterion (2) (relative errors 99.7% and 73% respectively), are achieved with the following parameters:

- NARX neural networks: start number K = 3; neuron number in layers N1 = 20, N2 = 10, N3 = 1; activation

characteristics of the hidden layers sigmoid, output layer linear; control and input TDL length: m = 10, n = 15, respectively.

- TDNN neural network: start number K = 3; neuron number in layers N1 = 5, N2 = 5, N3 = 1; activation characteristics of the hidden layers the sigmoid, output layer linear; input TDL length r = 5.

B. Simulation of Closed-loop Control System

Training closed-loop control systems suggests simultaneous adjustment of the object model and the control system parameters. Experiments were performed with architectural parameters of multilayer perceptrons including the NARX neural network and TDNN, as well as to the TDL lengths m, n, r. The best modeling precision values (Criterion (2)) on a test set are $I^{train} = 0.00052$, $I^{test} = 0.03709$, relative errors 99.9% and 84.4% respectively. Provided values are achieved on the following architectural parameters:

- NARX neural networks: K = 3, N1 = 5, N2 = 5, N3 = 1; m = 9, n = 8; activation characteristics of the network hidden layers sigmoid, output layer linear,

- TDNN: K = 3, N1 = 5, N2 = 5, N3 = 1; r = 6, activation characteristics of all the hidden network layers sigmoid, output layer linear.

Training of all the neural networks was carried out first by gradient descent method with momentum for $\tau = 70$ epochs, and then by Levenberg-Marquardt method for another $\tau = 500$ epochs. The need of using gradient descent method with momentum is associated with characteristics of adaptive-trained neural network topography, as well as the fact that the Levenberg-Marquardt method is ineffective in the search area, remote from local minima, where the network parameters are likely to fall as a result of their random initialization.

V. CONCLUSION

The conducted research confirmed the operability of the proposed blood glucose control system for IDDM patients, consisting of a set of connected neural networks of different architectures: NARX NN and TDNN. Despite achieved results, the proposed neural network model requires additional research, as well as determining the conditions of its applicability.

ACKNOWLEDGMENT

The research is supported by Russian Fundamental Research Fund (RFFI) grant No 10-08-00816-a.

REFERENCES

[1] S.A. Chernetsov, A.P. Karpenko, A.O. Emelianov, S.M. Gomenyuk , "Optimal Insulin Dose Forecasting Methods and Systems for Insulin-Dependant Diabetes Patients. Review" http://technomag.edu.ru/doc/119663.html, No 4, 2009.

[2] C. Osovski, "Neural networks for information processing.' - *Moscow: Finance and Statistics*, 2002.

[3] N. Todorovic "State of the art in nonlinear dynamical system identification using artificial neural networks" *Proceedings of 8th Seminar on Neural.*

[4] K. Narendra, K. Parthasarathy "Identification and control of dynamical systems using neural networks" *IEEE Transactions on Neural Networks*. –1990. – Vol. 1, № 1. – Pp. 4–27.

[5] B. Fernandez, A. Parlos, W. Tsai "Nonlinear dynamic system identification using artificial neural networks". *Proceedings of the International Joint Conference on Neural Networks.* – 1990. – Pp. 133-141.

[6] S. Chen, S. Billings, P. Grant "Non-linear system identification using neural networks" *International Journal of Control*. – 1990. – Vol. 51, № 6. –pp 1191-1214.

[7] S. Omatu, M. Khalid, R. Yusof, "NeuroControl and its Applications." *Moscow*: 2000.

[8] A.I. Galushkin "Neural networks theory" *Springer*, 2007.

[9] S. Haykin "Neural networks: a complete course". *Moscow: Publishing House "Williams"* 2006.

[10] A.P. Karpenko, K.N. Kosorukov, A.A. Saburov, S.A. Chernetsov, "Neural network-based blood glucose level forecasting in type I diabetes patients*". http://technomag.edu.ru/doc/1987.html, № 7*, 2011.

[11] S. Chernetsov, A. Emelyanov, A. Karpenko "Research of neural network-based blood glucose level forecasting systems for insulin-dependant diabetes patients", *6th International Workshop on Wearable Micro and Nano Technologies for Personalized Health (pHealth).* [*http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5754820*], 2009. Pp.29-32.

[12] H. Siegelmann., B. Horne, C. Giles "Computational capabilities of recurrent NARX neural networks : technical report ",*UMIACS* 1995.

[13] C. Alippi, "Experimental neural networks for prediction and identification" *Instrumentation and Measurement, IEEE Transactions* on. - 1996. - Vol. 45. - Pp. 670-676.

[14] K. Lang, A. Waibel, G. Hinton. "A time-delay neural network architecture for isolated word recognition", *Neural Networks.* – 1990. – № 3. – Pp. 23-43.

[15] A. G Trofimov, I. S Povidalo, S.A. Chernetsov "Application of self-organizing neural networks for identification of blood glucose level of 1-st type diabetic patients" *http://technomag.edu.ru/en/doc/142908.html* #10.

[16] Chernetsov S., A., Moor D., A. "Neuro-fuzzy prediction of the glucose level of patients with insulin-dependent diabetes" *http://technomag.edu.ru/en/doc/162907.html* #11.

[17] Chernetsov S.,A., Chuchueva I., A. "Blood glucose level forecasting in type I diabetes patients by neural networks and extrapolation by maximum similarity method" *http://technomag.edu.ru/doc/162847.html* #11.